REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
		FINAL - 30 SEP 92 TO 29 NOV 95

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
RESEARCH IN APPEARANCE DESCRIPTION FOR MACHINE VISION	F49620-92 -C-0073 8875/00

6. AUTHOR(S)

STEVE SHAFER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

CARNEGIE MELLON UNIVERSITY
SCHOOL OF COMPUTER SCIENCE
PITTSBURGH, PA 15213

AFOSR-TR-96

0185

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFOSR/NM 110 DUNCAN AVE, SUITE B115 BOLLING AFB DC 20332-0001	F49620-92-C-0073

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITIED.	

13. ABSTRACT (Maximum 200 words)

SEE REPORT FOR ABSTRACT

19960502 052

14. SUBJECT TERMS	15. NUMBER OF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

Standard Form 298 (Rev 2-89)

Final Report

ARPA Order: 8875

Program Code: 2E20

Contractor: Carnegie Mellon University

Effective Date of Contract: 30 September 1992

Contract Expiration Date: 30 September 1995

Amount of Contract Dollars: $199,577

Contract Number: F49620-92-C-0073

Principal Investigator: Steve Shafer 412-268-2527

Program Manager: Abe Waksman

Short Title of Work: **Appearance Description**

31 January 1996

1. Report Summary

1.1. Summary of Proposed Research

The key barrier to application of machine vision in unconstrained environments is the complexity of image formation in the world and the resultant difficulty of characterizing it concisely. If we could create a general yet concise description of image formation, we would have a vocabulary for discussing the complexity of specific scenes and the assumptions of specific machine vision approaches. In this research, the investigators are attempting to develop such a "vocabulary" consisting of a mathematical formalism for describing scenes, and examples of programs that utilize this formalism and data that is described using it. The data collection is not only for the purpose of this research contract, but also as a way of archiving and broadcasting high-quality image data from the Calibrated Imaging Laboratory at CMU, that may be useful for other researchers in image understanding.

1.2. Technical Results Summary

In the past three years of research, we have made tremendous progress towards a totally new concept for image segmentation based on the consideration of optical physics, rather than the heuristic clustering methods of the past. Our concept for the new algorithm is as follows:

1. Partition the image into "Uniform Chromaticity Regions", UCRs, of pixels with a uniform chromaticity but possibly varying intensity. These are supposed to approximate "appearance patches" of uniform physical explanation.

2. For each UCR, enumerate the plausible hypotheses about the formation of that region of pixel values.

3. Now, considering regions in pairs, keep only those pairs of hypotheses that provide the simplest explanation of that pair of UCRs.

Our work in the first year consisted of defining a reasoning framework we call the "taxonomy" of appearance elements; formalizing the concept of UCRs; enumerating the most plausible hypotheses for a single UCR based on the appearance taxonomy; and identifying a preliminary filtering criteria for hypotheses of adjacent regions. Taken together, these formed the backbone of a new algorithm for physics-based image segmentation.

During the second year, our principal accomplishments were the beginning of implementation of our method based on a representation of hypotheses, the segmentation of images into UCRs, and a systematic method for testing compatibility of hypotheses; and the production of community resources in the form of calibrated datasets and the Computer Vision Home Page. In addition, we now have a growing number of publications about this work.

One focus of our work in this reporting period has been on the initial partitioning of the image

into Uniform Chromaticity Regions (UCRs), which will be suitable building blocks for the reasoning process as the segmentation progresses. We have adopted normalized color as a basis for grouping, with a raster-scan region-growing algorithm.

We also developed a more systematic analysis of what it means for hypotheses to be "compatible". We defined that to mean that the pair of hypotheses must have all identical elements, except that a discontinuity exists in exactly one element of the hypothesis. Based on this definition, we created a table that tells how to make such a test for every candidate pair of hypotheses.

During 1994 we also generated our first calibrated dataset we called CIL-0001. As per a discussion with various ARPA IU Investigators and Oscar Firschein, Program Manager for IU at ARPA, we decided to collect a stereo/motion dataset rather than a color dataset as our first exercise. This dataset is therefore not so directly related to what is now the focus of this contract, which is color image segmentation. However, it was felt that a stereo/motion dataset would be far more valuable to the many researchers in those topics under ARPA sponsorship, and thus the dataset is presented as a service to the ARPA IU community. We have since added two more datasets and the Computer Vision Home Page.

During 1994 we published a technical report [2] and a conference paper[3] on the work in image segmentation. We also prepared and submitted a new paper to the ICCV-95 conference [4].

In the past year, 1995, our principal achievement is the implementation of the segmentation framework we outlined. This implementation focused on the most important pair of hypotheses we identified, piece-wise uniform dielectric objects under white illumination.

We tested both direct and implicit methods of analyzing adjacent hypotheses for compatibility, and found three implicit methods that provided a robust and effective measure. We now use two physical characteristics, the reflectance ratio and the direction of the gradient of image intensity, and an analysis of the intensity profile using information theoretic criteria. These three tests allow us to test the compatibility of two hypothesis regions more quickly, robustly, and in more complex scenes than direct instantiation techniques such as shape-from-shading and illuminant direction analysis.

We also showed how to combine the results of our three compatibility tests to generate a hypothesis graph, which reflects both adjacency in the image and the compatibility of hypotheses. From this hypothesis graph, it is possible to extract the best segmentation(s) of the image.

Our most recent work involves extracting segmentations from the hypothesis graph and integrating the components developed over the last several years into a complete segmentation system. By modifying a step-wise optimal merging algorithm to work on a multi-layer graph, we are able to extract the "best" segmentations of an image from the hypothesis graph, which contains information about the compatibility of adjacent hypotheses. By integrating this into our segmentation framework we now have a system to provide intelligent segmentations of images containing multi-colored objects. Furthermore, the framework is easily expandable to include more hypotheses, which increases the complexity of images the system can successfully and intelligently segment.

During 1995 we prepared and submitted four papers. The first is a technical report, detailing the compatibility tests and generation of the hypothesis graph [5]. A short version of this report appeared in the 1995 IUW proceedings [6]. The second, to appear in the 1996 IEEE Conference on Computer Vision One, is based on the technical report, but updated to show the final segmentations extracted from the hypothesis graph [7]. The third is a journal article submitted to Computer Vision and Image Understanding describing our entire theory and results to date [9]. The last, submitted to the Int'l Workshop on Object Recognition for Computer Vision demonstrates the necessity for a segmentation system that identifies objects, or coherent surfaces in a scene in order to obtain accurate object models from single images. We show that our framework provides such a segmentation [8].

1.3. Implications for Further Research

We have presented a framework for segmentation of complex scenes using multiple physical hypotheses for simple image regions. A consequence of this framework was a proposal for a new approach to the segmentation of complex scenes into regions corresponding to coherent surfaces rather than merely regions of similar color. Our work has progressed to an implementation of this new approach and we have shown example segmentations of scenes containing multi-colored piece-wise uniform objects. By using this new approach we are able to intelligently segment scenes with objects of greater complexity than previous physics-based segmentation algorithms. Our results show that by using general physical models we can obtain segmentations that correspond more closely to objects in a scene than segmentations found using only color.

These results have implications in model acquisition, object recognition, and the general analysis of color images. Furthermore, our framework is easily expandable. By incorporating more tools for analysis and more hypotheses, we can expand the range and complexity of images we can intelligently segment. This future work is fundamental to achieving effective image understanding and scene analysis.

2. Technical Results

2.1. A Taxonomy of Elements of Image Formation

In the proposal for this research, we presented a new approach to describing appearance elements -- the shape of the surface, its optical properties, and the incident illumination -- using functional notation. The general functions we presented are useful from a theoretical point of view, but they are not practical for reasoning about scene interpretation because they are "too precise". Instead of exact quantitative descriptions of the appearance elements, for scene interpretation we would be better off with overall categories such as "plastic", "metal", "diffuse illumination", "rough surface", etc. Such categories would correspond more closely to the human experience of vision, and to recognize them in general images would be a noteworthy achievement.

What would be most desirable is to have, for each of the scene elements, a categorization into an

ordered set of categories, from simplest to most complex. Then, when presented with an image to interpret, we could seek the simplest overall combinations of appearance elements to explain that image. Of course, this begs the question of what we mean by "simplicity". Since we lack any basis for answering this on physics grounds, we will merely appeal to intuition in developing and ordering our categories. Our experiments, later in this research program, will tell us whether we need to refine our category structure.

2.1.1. Categories for Surface Shape

A natural set of categories for surface shape is to begin by classifying surfaces as to whether they are curved, and if so, what degree of curvature they exhibit. Our approach is to classify surfaces according to the number of non-zero principal curvatures they exhibit, and further identify the case of identical curvatures, as follows:

Planar: Surfaces with zero curvature.

Cylindrical: Surfaces with one non-zero principal curvature.

Spherical: Surfaces with two non-zero principal curvatures, identical in value.

General Curved: Surfaces with two non-zero principal curvatures.

Initially, we will further simplify by grouping all the non-planar surface categories into a single category. In the case of surface shape, unlike the properties described below, there is little need to appeal to the underlying formalism of the definition of surface in our notation, that is, as an embedding $S(u,v) \rightarrow (x,y,z)$ mapping two-dimensional coordinates (u,v) to three-dimensional coordinates (x,y,z) over a subset E of the u-v manifold.

2.1.2. Categories for Incident Illumination

We categorize the illumination by reference to the incident light energy field, in our notation $L^+(x,y,z,\theta_x,\theta_y,\lambda,s,t)$ representing the amount of energy incident at point (x,y,z) from direction (θ_x,θ_y) with wavelength λ at Stokes (polarization) parameter s at time t. We begin by assuming no changes over time, so we can simplify to $L^+(x,y,z,\theta_x,\theta_y,\lambda,s)$. According to Figure 1, we then have several subcategories representing different simplifications of the light energy field.

2.1.3. Categories for Surface Optical Properties

We also categorize the surface optical properties based on our proposed transfer function, $T(x,y,z,\theta_x^+,\theta_y^+,\lambda^+,s^+,\theta_x^-,\theta_y^-,\lambda^-,s^-,t)$, which describes the distribution of all exitant light from a point invoked from a unit of incident illumination in any direction, wavelength, and polarization state. By assuming non-fluorescent material, no polarization selectivity, and no time-variance, this can be simplified at a point (x,y,z) to $T(\theta_x^+,\theta_y^+,\theta_x^-,\theta_y^-,\lambda)$, which is a familiar form of the Spectral Bi-Directional Reflectance Distribution Function [10]. As shown in Figure 2, we define further categories depending on the surface roughness, diffuse reflection distribution, and color of the specular and diffuse reflection.
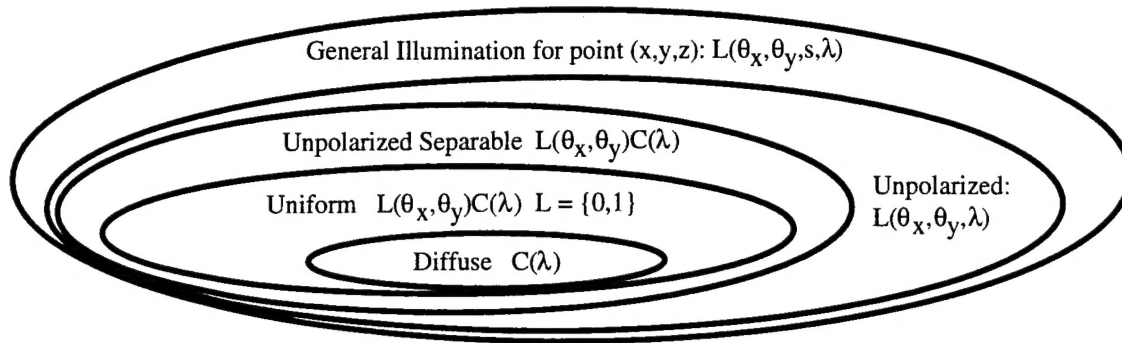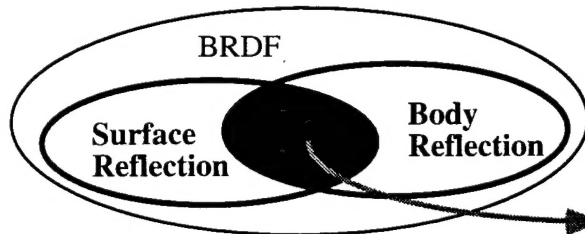
General Illumination for point (x,y,z): $L(\theta_x, \theta_y, s, \lambda)$

Unpolarized Separable $L(\theta_x, \theta_y)C(\lambda)$

Uniform $L(\theta_x, \theta_y)C(\lambda)$ $L = \{0,1\}$

Diffuse $C(\lambda)$

Unpolarized: $L(\theta_x, \theta_y, \lambda)$

Figure 1: Categories of the incident light energy

Spectral Bi-directional Reflectance Distribution Function
$$T(\theta_x^+, \theta_y^+, \theta_x^-, \theta_y^-, \lambda)$$

BRDF

Surface Reflection

Body Reflection

Special Cases of BRDF

White BR

NIR Reflection

Gaussian Facets   Non-Gaussian

Lambertian   Non-Lambertian

Figure 2: Taxonomy of Spectral BRDFs

2.2. Appearance Patches

Our new concept for image segmentation begins with the notion of an *appearance patch*, by which we mean a surface in the scene, or a piece of a surface, which exhibits coherence in the illumination, shape, and optical properties, i.e. in all three elements of our appearance description formalism. It is our goal to identify such appearance patches in the image, and to ascribe to them one or more most plausible explanations. The appearance patch is a fundamental concept because each such patch ought to have a single explanation, and the boundaries of the patch are the boundaries of applicability of that explanation.

At this point, we are careful not to be too strict in defining *coherence*, but simply to say it is some kind of uniformity, structure, or statistical regularity in the nature of the appearance description functions. Future research will be needed to determine exactly what constitutes useful coherence.

We have already defined a hypothesis, in our proposal, as a tuple of instantiations of the

appearance description functions for the illumination, object shape, and optical properties. Now, we add the concept of a *hypothesis set*, which is the set of hypotheses currently under consideration to explain a single appearance patch. An appearance patch, together with its attendant hypotheses, we call a hypothesis region.

In the course of processing, we imagine three steps:

4. Identify appearance patches in the image.

5. For each one, propose all plausible hypotheses to form a hypothesis set.

6. By looking at adjacent regions, identify compatible and incompatible hypotheses. Through this process, hopefully most hypotheses can be rejected so that only one or a small number remain at each region.

All three of these steps will require further research, and indeed, we don't even know at this time if they are possible to achieve.

And now, we are in a position to define what we mean by a *segmentation*: A segmentation is a set of hypothesis regions, each containing a single hypothesis, which are consistent with each other, and which explain all the regions of the image. In other words, a segmentation consists of a set of hypotheses which cover the image, providing a unique and plausible explanation of the illumination, shape, and material optics at each pixel. The segmentation process may produce one or more such segmentations, because in many cases there will be ambiguities that cannot be resolved by low-level vision. Still, this definition of segmentation is a major step forward in the science of image understanding. It is based on essentially objective criteria, and gives a much richer description of the scene and the objects in the scene than has been proposed by any segmentation program in the past.

2.3. Uniform Chromaticity Regions

We define a *uniform chromaticity region* [UCR] to be a connected set of pixels that possess uniform chromaticity and possibly varying brightness (Figure 3). A UCR corresponds to a linear cluster, as defined by Klinker *et al.*[12]. As such, a more general definition of a UCR is a connected set of pixels whose covariance matrix in color space has a single non-zero eigenvalue, whose eigenvector is related to the chromaticity of the region. Because it allows for varying brightness within a region, a UCR is able to capture more of the relevant coherence between
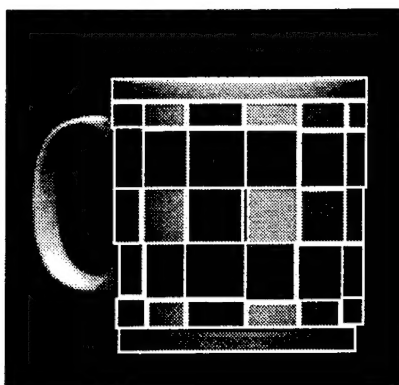
neighboring pixels than simple uniform regions.

Figure 3: Mug divided into idealized UCRs (by hand)

Klinker *et al.* [11] note that a UCR, or linear cluster, can represent two distinct objects if both are dark or poorly illuminated. In our segmentation method, however, we initially assume that a UCR represents a single surface patch under a single illumination environment. This requires a form of coherence from the physical elements generating the UCR. Clearly, it is possible to construct an image with UCRs that do not have such coherence in the physical world, and we realize that our current approach will not correctly handle such situations.

The benefit derived by using UCRs is that they are groupings of pixels that we can reasonably assume to correspond to a single appearance patch in the physical world, setting constraints on the associated hypotheses. These constraints are that over the patch the transfer functions are coherent and the illumination environments are similar. Because it is a single appearance patch, it is, by definition, a single surface. Figure 3 shows an idealization of the cup image divided into UCRs.

By identifying UCRs in the image, we have taken the first step in the segmentation process by linking pixels with appearance patches in the scene. The next step is to begin to identify the relevant physical explanations, or hypotheses, for the appearance patches corresponding to the identified UCRs.

2.4. Fundamental Hypotheses

An explanation for the color of a physical appearance patch of uniform chromaticity can be described in terms of several basic properties: the illumination environment, the material (body reflection and surface reflection), and the color source(s). Given a set of fundamental values of illumination, material, and color source, a finite list of hypotheses can be derived giving multiple explanations for this single UCR. This basic list consists of 42 fundamental hypotheses, each of which can explain the color of the patch.

Upon closer examination of these 42, it is clear that not all hypotheses are equally likely in the real world. For example, attributing the color of a green patch to a green material under a white light source is more common than attributing the color to a white material under a green light
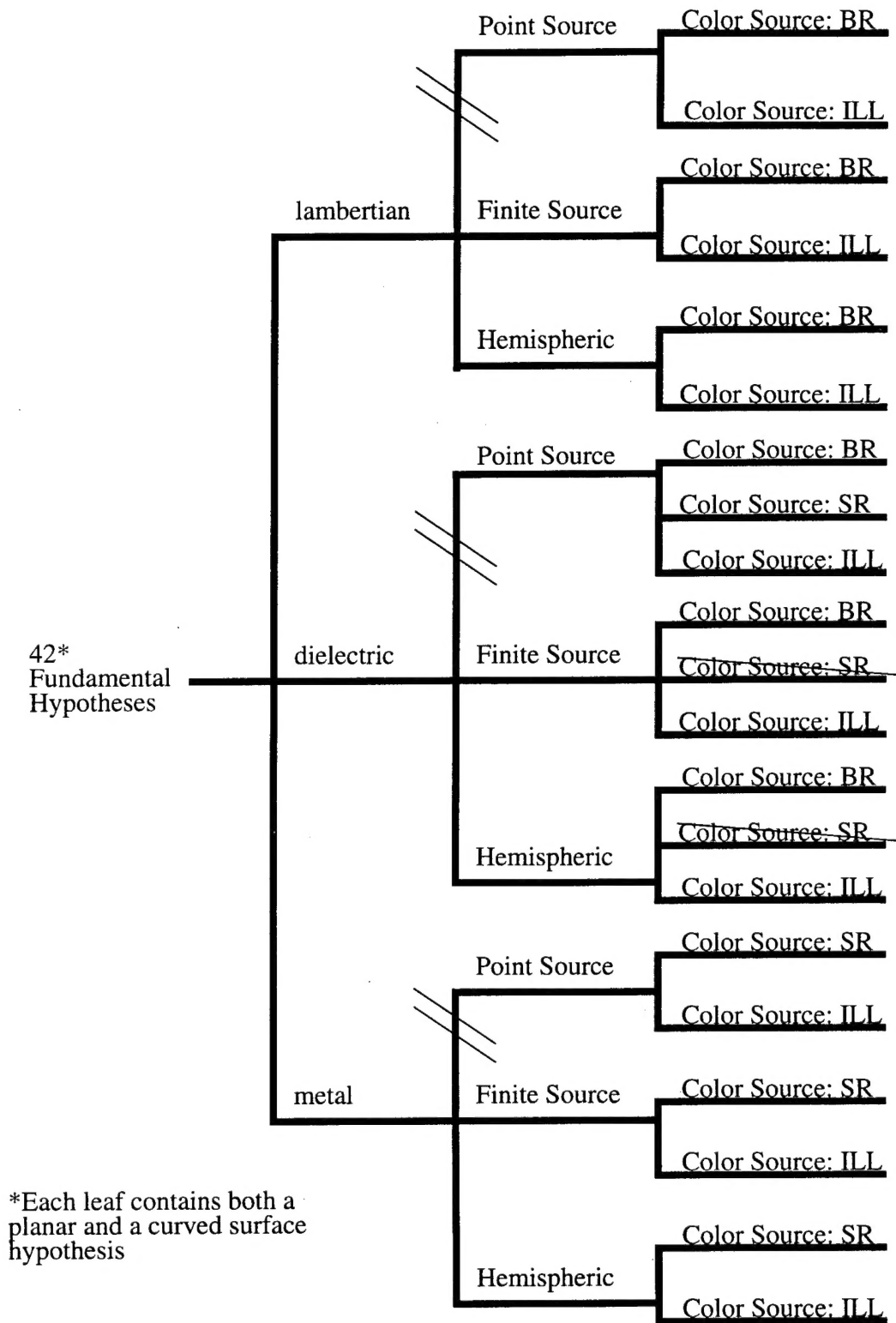
Figure 4: Taxonomy of Appearance Patch Hypotheses

The tree structure contains:

- **42* Fundamental Hypotheses**
  - **lambertian**
    - Point Source
      - Color Source: BR
      - Color Source: ILL
    - Finite Source
      - Color Source: BR
      - Color Source: ILL
    - Hemispheric
      - Color Source: BR
      - Color Source: ILL
  - **dielectric**
    - Point Source
      - Color Source: BR
      - Color Source: SR
      - Color Source: ILL
    - Finite Source
      - Color Source: BR
      - Color Source: SR
      - Color Source: ILL
    - Hemispheric
      - Color Source: BR
      - Color Source: SR
      - Color Source: ILL
  - **metal**
    - Point Source
      - Color Source: SR
      - Color Source: ILL
    - Finite Source
      - Color Source: SR
      - Color Source: ILL
    - Hemispheric
      - Color Source: SR
      - Color Source: ILL

*Each leaf contains both a planar and a curved surface hypothesis

source. In order to achieve a more structured ordering of the hypotheses, a tree structure was developed, producing a taxonomy of physical appearance. This taxonomy is given in Figure 4.

The first branching in the taxonomy deals with material type, specifically, whether the material is Lambertian (just body reflection), dielectric (body reflection and surface reflection), or metal (surface reflection only). In the general case, it is not possible to prune the tree here, as all three material types exist. If, however, a vision system is working in a limited environment, it may be possible to set probabilities for each branch or eliminate one of the branches altogether.

The second branching depends upon the illumination environment. If synthetic images (for example, images created by a ray tracing program) are considered, it is not possible to prune the tree here. If, however, the task is taking place within the real world, the point source branch can be pruned, as it is highly unlikely that such a source would naturally exist. This would reduce the number of total hypotheses from 42 to 28.

The third branching determines the color source of the patch. In the case of the Lambertian and metal surfaces, no pruning can be undertaken. It is highly unlikely, however, that a dielectric would have a uniform body reflection and a colored surface reflection in any domain except synthetic images. This allows another four hypotheses to be pruned, leaving 24 fundamental hypotheses for real world applications.

Besides providing for an orderly pruning of the list of fundamental hypotheses, this taxonomy is also useful for:

- Benchmarking and classifying existing and new physics-based vision programs.

- Indicating when different physics-based vision techniques are applicable.

- Developing scene descriptions for a data base.

- Pointing out unfulfilled needs in physics-based vision and indicating a future agenda for research.

2.5. Relations Among Adjacent Regions

With our concept of Fundamental Hypotheses, we have the outline of a segmentation algorithm: (1) partition the image into appearance patches of uniform color; (2) assign all Fundamental Hypotheses to each patch; and (3) look for sets of hypotheses that can explain groups of patches in the image. As a first step towards #3, we have developed a table (Table 1) that shows which

hypotheses might be compatible with which other hypotheses at an adjacent region.

|	diffuse		col. diel. uniform		grey diel. general ill.		col. metal uniform ill.		grey metal general ill.	
	planar	curved	planar	curved	planar	curved	planar	curved	planar	curved
diffuse planar	■		■							
diffuse curved		■		■						
col. diel. uniform planar	■		■							
col. diel. uniform curved		■		■						
grey diel. general ill. planar					■					
grey diel. general ill. curved						■				
col. metal uniform ill. planar										
col. metal uniform ill. curved										
grey metal general ill. planar									■	
grey metal general ill. curved										■

■ = desired merger

Table 1: Compatible Hypotheses for Adjacent Regions

What we mean by "compatible" is itself a subject of some study. Suppose we have a hypothesis H1 at one region, and hypothesis H2 at an adjacent region. We can always form hypothesis H1+H2 to describe the pair of regions. This new hypothesis describes a shape which has two independent parts; material optics that differ from one region to the other; and an illumination environment that has two distinct natures over the two regions. This would be completely uninteresting.

On the other hand, suppose that both hypotheses H1 and H2 had the same shape, and the same material optics, and that they differed only in that one assumed a bright light source and the other did not. This would be tremendously significant; the hypothesis H1+H2 would be just a tiny bit more complex than either of its elements (but far less complex than the two elements taken separately), and it might be interpreted as a "shadow falling across a single surface". Similarly, if H1 and H2 have the same shape and illumination, they might represent a single surface with different colored regions. So, what we seek are pairs of hypotheses such that explaining them together is simpler than explaining the two of them separately.

Our compatibility table was based on several criteria that try to capture what we mean by having a single surface composed of several appearance patches:

- Hypotheses of differing materials should not be combined.

- Hypotheses of differing color sources should not be merged.

- Hypotheses of differing shape should not be merged.

- "Colored metal" hypotheses of differing chromaticity and similar illumination should not be merged.

- If the hypotheses differ in their chromaticity and the illumination is the color source, then hypotheses with diffuse illumination environments should not be merged.

This table of compatibility may not capture all of the criteria needed for effective image segmentation. But, the fact that it is fairly sparse gives us hope that it will in fact yield a dramatic reduction in the number of hypotheses to be considered as possible image segmentations.

2.6. Illustrating a Hypothesis

Our "hypotheses" for image region interpretation consist of three elements:

- A shape function, which maps 2D surface coordinates to 3D world coordinates

- An illumination environment, which maps the incident light onto a hemisphere around the surface

- A transfer function, which describes how the incident light is transformed into exitant light

We have a new way of depicting these elements of a hypothesis, by means of several small image elements:

- The shape function is shown by a wire-frame representing a grid on the surface

- The illumination environment is depicted by a disk representing a perpendicular projection of the illumination hemisphere onto the plane of the surface

- The transfer function is shown by a little image of a sphere with the given transfer function, in an environment consisting of a checkerboard beneath the sphere, a black "sky" above, with a white point light source in the "sky" above and behind the camera's view.

These elements are depicted in Figures 30-39 of the technical report [2]. All the elements work well except that the sphere used to show the transfer function is a rather simplified view of the true complexity of the transfer function. However, it does serve to distinguish the most important cases (colored v. white, metal v. dielectric, smooth v. rough).

2.7. Example Worked by Hand

The technical report also includes our first attempt to formalize our reasoning framework by working through a complete example by hand. The image in this example is simply a sphere containing three colored patches -- two green areas with a blue stripe separating them. The goal is to show how a system could use our hypothesis framework in a general way to reason that the correct interpretation is that of a single shape element and illumination environment, with three differently colored patches of the same material type.

In our example, we consider only the ten best Fundamental Hypotheses for each region, which include all the most common situations. This would naively yield 100 hypothesis combinations for two regions, or 1000 for the three taken together. However, with our reasoning framework for pairwise interpretation of regions, described in an earlier report, we find only twelve combinations for each pair of regions, or 20 for all three taken together. This reduction by a factor of 50 is a strong vindication of our reasoning approach. Furthermore, by applying some reasonable heuristics to this set of possible interpretations, they can be grouped into rough categories from most likely to least likely. In our grouping, the only combined hypothesis in the top category is in fact the most likely one.

Table 2: . Final set of hypotheses for the example image

Hypothesis	Top Region	Middle Region	Bottom Region
1 Tier 1	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Uni./Curved
2 Tier 2	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Dif./Curved
3	Diel/CS=BR/Uni/Planar	Diel/CS=BR/Uni/Planar	Diel/CS=BR/Uni/Planar
4	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Dif./Planar
5	Metal/CS=IL/gf/Curved	Metal/CS=IL/gf/Curved	Metal/CS=IL/gf/Curved
6	Metal/CS=IL/gf/Planar	Metal/CS=IL/gf/Planar	Metal/CS=IL/gf/Planar
7 Tier 3	Diel/CS=IL/gf/Curved	Diel/CS=IL/gf/Curved	Diel/CS=IL/gf/Curved
8	Diel/CS=IL/gf/Planar	Diel/CS=IL/gf/Planar	Diel/CS=IL/gf/Planar
9 Tier 4	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Diff/Curved
10	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Uni./Curved
11	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Dif./Curved
12	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Uni./Curved
13	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Diff/Curved
14	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Uni./Curved
15	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Diff/Planar
16	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Uni./Planar
17	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Dif./Planar

Table 2: . Final set of hypotheses for the example image

Hypothesis	Top Region	Middle Region	Bottom Region
18	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Uni./Planar
19	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Diff/Planar
20	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Uni./Planar

The attached technical report [2] illustrates and explains the process in more detail.

2.8. Representation of Hypotheses

To represent hypotheses in the computer, we need data structures that not only capture the parameterization of each element of a hypothesis, but also allow hypotheses to be tested for compatibility and possibly merged into larger hypotheses.

For a surface patch, if planar, we simply represent a point on the plane and a direction vector for the surface normal. For a curved patch we use a 4x4 Bezier patch; these are easy to fit to data, and at the edges they line up along cubic Bezier curves, thus facilitating merger into larger hypotheses. One problem with the Bezier patches is than when merged, that result is not another Bezier patch, so this leads to a structure of Bezier patches rather than a single ever-growing patch.

For illumination hypotheses, in simple parameterized cases such as totally diffuse light, we just store the parameters. For the most complex case, general illumination, we use a map of the color at each direction. For intermediate cases, which are perhaps the most interesting, we store a list of extents and colors. Any of these representations allows easy merger with another hypothesis of the same type; combinations can get messy and would require a "coercion" of data into the more general type.

Similarly, for the transfer function, in a simple case such as a metal or Lambertian dielectric surface, we store just the parameters; but for more complex cases we can store a list of extents and colors or even a map of the reflectance properties at every point.

These representations are all summarized in Figure 5. The implementation is in C++ on a Macintosh Quadra.

Hypothesis

- Shape
    - enumerated type s ∈ {planar,curved}
    - plane type p = list of (point,normal,extent)
    - curve type c = list of (4x4 bezier curve,extent)
    - mapping function $M(u, v) \rightarrow (x, y, z)$)
    - inverse function $M^{-1}(x, y, z) \rightarrow (u, v)$
- Illumination
    - enumerated type l ∈ {diffuse,uniform,general,mixed}
    - diffuse type d = (color,extent)
    - uniform type u = (color,bitmap,extent)
    - general type g = (pixel map,extent)
    - mixed type m = list of
        - diffuse type l
        - uniform type l
        - general type
- Transfer function
    - enumerated type r ∈ {metal,dielectric}
    - metal type m = (color,roughness)
    - dielectric type d ∈
        - uniform = (color,roughness,extent) l
        - piecewise = list of (color,roughness,extent) l
        - general = (pixmap)

Figure 5: Representation of a Hypothesis

2.9. Finding UCRs

We define a UCR to be a region of constant chromaticity (hue and saturation), but possibly varying intensity of pixel values. UCRs are important because they capture what is important in our reasoning process, and eliminate what is invariant.

What is important in our reasoning process is to identify points with the same combination of illumination and surface coloration, so that these points can be grouped into a single unit with a single hypothesis set to explain them as a group. Their adjacency to other such groups, with clearly distinct coloration of the illumination or surface, is critical to the reasoning process we will apply. At the same time, we must not build into the grouping any assumption about the actual color of the illuminant or objects, because we know that these may vary from image to image or even within a single image (particularly the illuminant). The UCR accomplishes these purposes. It makes clear where the boundaries of coloration lie, without confounding them with shape boundaries (associated with intensity changes). Further, if the illuminant is not white, its chromaticity will interact with that of the UCR is in principle well suited for our use as our building-block regions in the image.

To determine UCRs, we abandon the usual R-G-B space and jump instead into normalized color

space, defined by dividing the R, G, and B values by their sum R+G+B. We threshold these values to decide whether the color is too close to black to be reliably classified on the basis of chromaticity.

We partition the image by making a single raster-scan over it. During this scan, if we detect a suitable seed patch for a new region, which is not yet part of a region, then we interrupt the raster scan and grow the newly discovered region. Then, having marked all the pixels in that new region, we resume the raster scan.

We consider a patch of pixels suitable for use as a seed region when its central pixel and its 8-connected neighbors have similar normalized color, and none of these pixels already are marked as belonging to another region. The growing process considers the 4-connected neighbors of all pixels in the region, and if they are close enough to the average normalized color of the seed patch, and not too dark to be reliable, then they are incorporated into the new region.

This method of segmentation does a nice job of finding regions to grow, but one problem is that it leaves rather ragged edges. Since we must know which regions are adjacent to which other regions, we have to deal somehow with these unlabeled pixels that separate the cleanly defined regions. That is a task for the future.

2.10. Analyzing Compatibility of Hypotheses

In the past, we did not have a very systematic definition for what it means for two hypotheses to be "compatible"; we only had a heuristic sense that they ought to have some elements in common. Now, we have a cleaner definition: We consider two hypotheses to be "compatible" iff they share all elements except one, which exhibits a discontinuity coincident with the region boundary. Note that such a criterion is actually much stricter than our previous table (Figure 6), which by its gray boxes merely indicates what *might be* compatible hypotheses. Further tests are needed to determine if specific hypotheses corresponding to those cases are indeed compatible by the above definition.
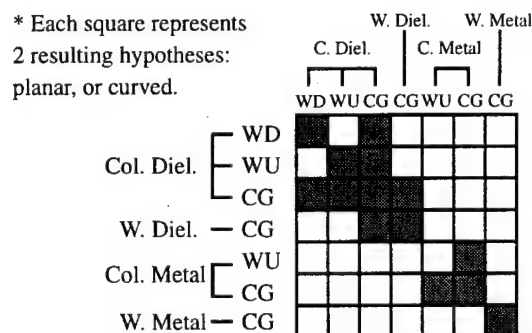
Figure 6: Possibly Compatible Fundamental Hypotheses

We have conducted a study of these cases, and we have a new table that shows how to conduct these tests in a systematic way (Table 3). The "Comments" in this table are actually not merely comments, but represent our insight so far about how one might go about constructing a program

to perform the required test. We are now preparing to undertake such a test for the most important single case, that of two dielectric surface patches of the same surface, with different colors.

Table 3: Analysis of Merger Table

Input Region 1	Input Region 2	Discontinuity	Comments
white diffuse/col. dielectric	white diffuse/col. dielectric	transfer function	self-shadowing, 2-D cues
white diffuse/col. dielectric	colored general/ col. dielectric	illumination	known object color restricts color of region 2, 2-D cues
white uniform/ col. dielectric	white uniform/ col. dielectric	transfer function	shape-from-shading, illuminant direction & color, roughness estimation
white uniform/ col. dielectric	colored general/ col. dielectric	illumination	shape-from-shading, known object color restricts color of region 2, color constancy, roughness estimation
colored general/ col. dielectric	colored general/ col. dielectric	illumination or transfer function	orientation-from-color, roughness, illuminant color & direction estimation
colored general/ col. dielectric	colored general/ white dielectric	transfer function	illuminant direction estimation, illuminant color known from region 2, roughness, orientation-from-color
colored general/ white dielectric	colored general/ white dielectric	illumination	roughness estimation, orientation-from-color
white uniform/ colored metal	colored general/ colored metal	illumination	roughness estimation, known metal color, 2-D cues
colored general/ colored metal	colored general/ colored metal	illumination	roughness estimation, estimation of metal color, 2-D cues
colored general/ white metal	colored general/ white metal	illumination	roughness estimation, 2-D cues

The significance of this table is that it shows where and how various methods for physics-based computer vision can be blended together to make a single system whose reasoning and interpretive power is far greater than any one method taken in isolation. The physics-based vision community has been working blindly for so many years, without a good problem statement to constrain their search for useful algorithms; we believe that this table represents such a constraint and provides a map for the definition of what algorithms will be "useful" to see in the future.

2.11. Stereo Image Datasets

One of the topics of greatest interest in the ARPA Image Understanding community is the determination of shape by analysis of stereo or motion image sets. Many such datasets exist, however they are almost always lacking in ground truth [13]. Therefore, there is no way to judge or measure the quality of the resulting interpretations by computer programs. In an effort to advance a more scientific approach, the CIL has used its unique facilities to collect a dataset with

careful calibration of both the imaging conditions and the scene itself.

The imaging facilities used for this data collection include a high-precision cooled CCD camera with very low geometric and radiometric noise, a high-precision motion jig capable of 0.01^o angular steps and 0.001in translation steps in all six degrees of freedom. The ground truth data (3D point coordinates) are collected by a pair of surveyor's theodolites with an estimated precision of about 0.3mm in each dimension.

The images are available by anonymous FTP, World Wide Web, and e-mail. The 11 images in the dataset were acquired as indicated in that document in an arrangement that allows them to be used simulate lateral stereo, vertical stereo, multi-camera arrays, and forward camera motion by one or two cameras.

The ground truth data consists of two elements. First, for each of the camera positions, images were taken of grids at varying distances, and a Tsai-Lenz camera calibration was performed to determine the extrinsic parameters (pose) and intrinsic parameters (focus etc.) of the camera. The results of these calibrations are presented in the dataset. Second, we measured the 3D position of 23 points with the surveyors' theodolites, and also located them by hand in each of the images. These 3D and 2D coordinates are also reported in the dataset. Whenever possible, these points include "triangle sets" of at least 3 points on a single flat surface, so that interpolated depth results can be evaluated across every pixel of the included triangle rather than just at the corner points themselves.

We hope and believe this dataset will be of fundamental value to the ARPA IU community (including our own work in the areas of stereo and motion). However, due to the extraordinary difficulty and cost of obtaining such data, we cannot realistically expect to acquire such datasets very frequently. A reasonable estimate would be one to two man-months to acquire this dataset, and while we have tried to automate the process as much as possible, the fact is that there remains a high degree of human action involved, and usually several attempts must be made before any particular aspect of the data collection effort can be pronounced successful.

2.12. The New Stereo Datasets

The new stereo datasets, CIL-0002 and CIL-0003, are similar to our CIL-0001 dataset, but differ in the nature of the scene. We have found that one of the key tasks in stereo is the reconstruction of large surface patches, which the CIL-0001 dataset did not facilitate. In CIL-0002, we have a single surface in the scene, so this dataset is much more useful for those people trying to study interpolation of depth and disparity. And, in CIL-0003, we have a combination in which there are a very small number of fairly large surfaces, so that the interpolation over large regions is combined with 3D shape modeling. By having a small number of surfaces, the ground truth points in CIL-0002 and CIL-0003 allow reconstruction of the interpolated depth values more easily than our previous dataset CIL-0001 which had a far more complex scene.

2.13. The Computer Vision Home Page

The "Computer Vision Home Page" is our own initiative to try to set up one centralized location with pointers to all aspects of image understanding. These include pointers to research group descriptions, pointers to individuals, lists of available datasets for algorithm testing, hardware sources, upcoming conferences, and executable demonstrations of vision programs and cameras. The success of such a page depends mostly on the completeness of its listings, so we are encouraging other sites to submit to us the information that will help flesh this out and make it a valuable community resource.

2.14. Compatibility Testing by Shape-From-Shading

Our table of compatible hypotheses does not tell when two hypotheses must be compatible; it only tells when they might be compatible. Therefore, while an empty box definitively rules out a combination, a gray box does not mean that this pair of hypotheses must always be accepted as compatible; it only means there is a possibility. A further test is needed for each such gray box, to tell whether this particular pair of hypotheses for this particular pair of regions, is actually compatible or not. Thus, the discriminative power of our framework (already 50:1 from the white squares in the table) will be considerably enhanced when we have added all of these additional tests to our program. Indeed, not only will there be fewer incorrect hypotheses being propagated, but we hope and believe that it will become hard for the system to ever maintain an incorrect hypothesis for a very long time.

We have begun to implement such tests by selecting the one most important gray box from the table and implementing a test for that one case: the case of two regions of colored dielectric material, both curved surfaces, under diffuse illumination. This is the situation describing, for example, a single curved surface with two colored patches side-by-side, such as a ball with a stripe on it, or a cup with a colored design on it, or a print fabric, or a piece of paper with a picture drawn on it.

For two such hypotheses to be considered "compatible", we already know that the transfer functions differ in color, but not in material type. So far, so good. But, what about the surface shape and the illumination environment? These must now be tested for compatibility.

For shape compatibility, the obvious solution is to apply a Shape-From-Shading method, of which there are over a dozen in the literature, to calculate the shape of each region. If these are the same along their common edge, then the regions are compatible based on shape. The biggest problem is that most methods for Shape-From-Shading begin by assuming that they will process the entire image, or at least a closed region surrounded by a tangency contour. In other words, they cannot process a piece of a surface; it is all or nothing. However, in our case, we must process pieces of surfaces, not necessarily surrounded by tangency boundaries. Therefore, we cannot utilize most of the methods that are published. Instead, we use a "local" method, based on some pointwise approximation to the surface shape. The best method appears to be that of Bichsel and Pentland [14], so this is what we use. After fitting the two region shapes, we look at the resulting common boundary. We must allow for affine scaling and translation in the z direction in assigning a

compatibility score.

For the illumination environment, both regions better indicate the same illumination situation. Since Shape-From-Shading methods generally assume point lighting, we adopt this view. Then, we need a method to estimate the angle of the light source based on the shading analysis. The best method we have found for this is that of Zheng and Chellappa [15]. So, we apply this to each region to determine the angular difference between the two computed light source directions. If they are close, we give the region a good compatibility score.

One thing we do not yet have a good idea for is how to combine the scores for different tests at a region. On the face of it, multiplying them would seem be appropriate. But, this would penalize hypothesis combinations for which lots of tests exist. Instead, we may decide to take a "worst-case result" of the set of compatibility results.

If we can estimate and represent each hypothesis element, merging adjacent regions involves looking at the table of possible mergers and then directly comparing the values of each hypothesis element. We attempted to implement the direct instantiation approach for the hypotheses (Colored plastic, White Uniform illumination, Curved) and (White plastic, White Uniform illumination, Curved) for which some tools of analysis do exist for finding both the shape and illumination of a scene.

Our conclusion was that the basic problem with the direct instantiation method is that it requires region-based analysis. Existing tools for analyzing the intrinsic characteristics of a scene cannot, in general, be used on small regions of an image because it violates basic assumptions necessary for the tools to function properly. Furthermore, if we attempt to generalize direct instantiation to other hypotheses or more complex situations, we are currently limited by the lack of image analysis tools. While other approaches to shape-from-shading may overcome some of these difficulties in the future, for now we take a different approach.

2.15. Compatibility by Implicit Instantiation

An alternative to direct instantiation of hypotheses is to use the knowledge constraints provided by the hypotheses to find physical characteristics that can differentiate between pairs of regions that are part of the same object and pairs of regions that are not. As these physical characteristics are generally local, they are more appropriate for region-based analysis than the previously mentioned direct-instantiation techniques. We call this method *implicit instantiation*.

2.15.1. Reflectance Ratio

One physical characteristic we use is the reflectance ratio for nearby pixels as defined by Nayar and Bolle [16]. The reflectance ratio is a measure of the difference in transfer function between two pixels that is invariant to illumination and shape so long as the latter two elements are similar. If the shape and illumination of two pixels p_1 and p_2 are similar, then the reflectance ratio, defined in equation (1), where I_1 and I_2 are the intensity values of pixels p_1 and p_2, reflects the change in albedo between the two pixels [16].

$$r = \left(\frac{I_1 - I_2}{I_1 + I_2}\right) \tag{1}$$

For each border pixel p_{1i} in h_1 that borders on h_2 we find the nearest pixel p_{2i} in h_2. If the regions belong to the same object, and therefore have similar shape and illumination but differing transfer functions, the reflectance ratio should be the same for all pixel pairs (p_{1i},p_{2i}) along the h_1,h_2 border. A simple measure of constancy is the variance of the reflectance ratio. If the two hypotheses being tested are part of the same object, this variance should be small, due mostly to the quantization of pixels, noise in the image, and small-scale texture in the scene. If, however, h_1 and h_2 are not part of the same object, then the illumination and shape are not guaranteed to be similar for each pixel pair, violating the specified conditions for the characteristic. Differing shape and illumination should result in a larger variance in the reflectance ratio. We select an expected variance based upon the noise, variance in object's transfer functions, and quantization effects and use this expected variance to differentiate between these two cases. Table 4 shows the results of this operator when applied to the image of a stop-sign and cup shown in Figure 7.

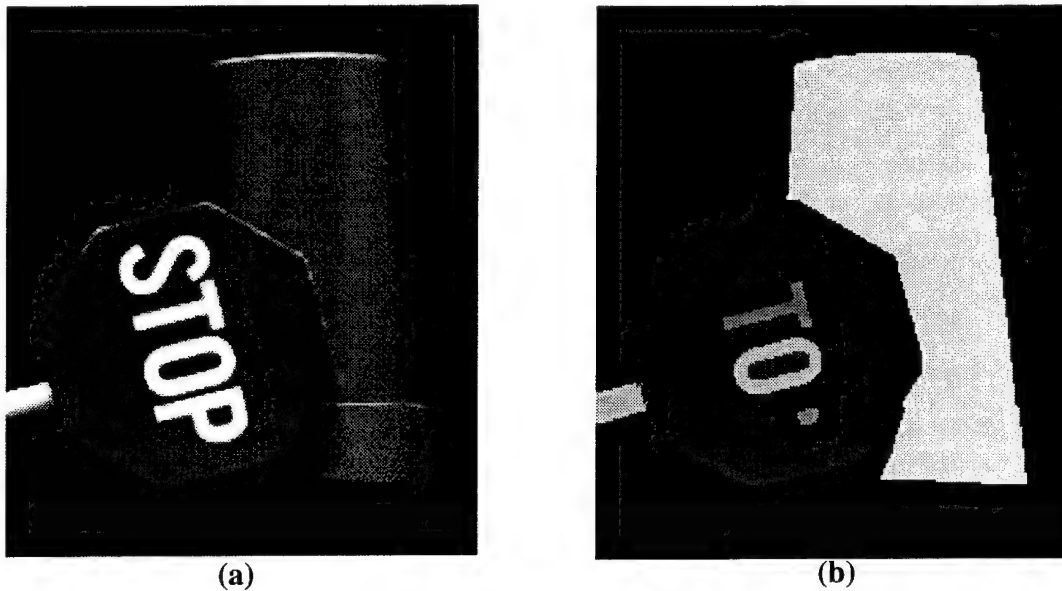(a)   (b)

Figure 7: (a) Image of a red and white stop-sign and a green cup taken in the Calibrated Imaging Laboratory, CMU. (b) initial segmentation of the image

Table 4: Reflectance ratio for stop-sign and cup image

Region 1	Region 2	μ (avg. RR)	σ^2	$P(\sigma^2 < \Sigma^2)$
Sign	Letter S	.4463	.0004	1.0
Sign	Letter T	.4449	.0005	1.0
Sign	Letter O	.4503	.0004	1.0
Sign	Letter P	.4541	.0006	1.0
Sign	Cup	.2107	.0125	0.0
Sign	Pole	.1709	.0710	0.0
Letter O	O hole	-.4358	.0008	1.0
Letter P	P hole	-.4562	.0004	1.0

2.15.2. Gradient Direction

The direction of the gradient of image intensity can be used in a similar manner to the reflectance ratio. The direction of the gradient is invariant to the transfer function for piece-wise uniform dielectric objects--except due to border effects at region boundaries. Therefore, by comparing the gradient direction of border pixel pairs for two adjacent regions we obtain an estimate of the similarity of the shape and illumination. To avoid border effects, the algorithm first calculates the gradient direction of non-border pixels and then grows the results to include border pixels.

As with the reflectance ratio, we sum the squared difference in the gradient directions of adjacent border pixels from two hypotheses to find the sample variance for each hypothesis pair. We then use this variance to differentiate hypotheses pairs that are likely to be part of the same surface from those that are not. Figure 8 shows a visualization of the gradient direction errors for the image of two-spheres.
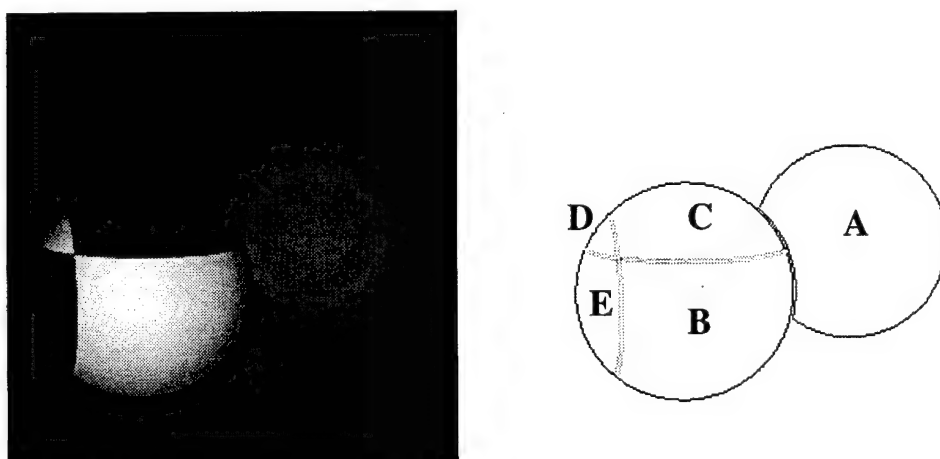

Figure 8: Result of gradient direction analysis on a synthetic image of two spheres. Darker borders indicate greater error.

2.15.3. Intensity Profile Analysis

So far, we have examined only calculated physical characteristics of the image, not the actual image intensities. The intensity profiles contain a significant amount of information, however, which we attempt to exploit with the following assertion: if two hypotheses are part of the same object and the illumination and shape match at the boundary of the hypotheses, then, if the scale change due to the albedo difference is taken into account, the intensity profile along a scanline crossing both hypotheses should be continuous. Furthermore, we should be able to effectively represent the intensity profile across both regions with a single model. If two hypotheses are not part of the same object, however, then the intensity profile along a scanline containing both hypotheses should be discontinuous and two models should be more appropriate to represent it.

To demonstrate this property, consider Figure 9(b), which shows the intensity profile for the scanline from A to A'. We can calculate the average reflectance ratio along the border to obtain the change in albedo between the two image regions. By multiplying the intensities from A" to A' by the average reflectance ratio we adjust for the difference in albedo. As a result, for this particular case the intensity profile becomes smooth. On the other hand, for the scanline B to B', shown in Figure 9(b) the curves are not smooth even when the intensities are adjusted.
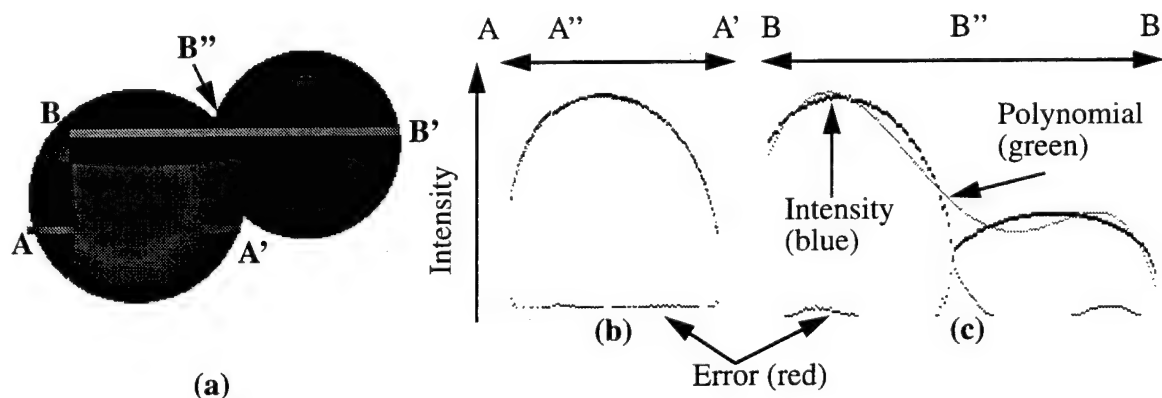

Figure 9: Test image shown in (a). Graphs (b) and (c) are the intensity curves and least-squares polynomial for the image segments A-A' and B-B', respectively.

Rather than use the first or second derivatives of the image intensities to find discontinuities, we take a more general approach which maximizes the amount of information used and is not as sensitive to noise and small-scale texture in the image. Our method is based upon the following idea: if two hypotheses are part of the same object then it should require less information to describe the intensity profile for both regions with a single model than to describe the regions individually using two. We use the Minimum Description Length [MDL], as defined by Rissanen [17], to measure complexity, and we use polynomials of up to order 5 to approximate the intensity profiles. The formula we use to calculate the description length of a polynomial model is given in equation (2), where x^n is the data, θ is the set of model parameters, k is the number of model

parameters, and n is the number of data points [17].

$$)L = -\log P(x^n|\theta) + \frac{k}{2}\log i \tag{2}$$

Our method for a single scanline is as follows.

7. Model the intensity profile on scanline s_0 for hypothesis h_1 as a polynomial. Use the MDL principle to find the best order polynomial (we stop looking after order 5). Assign M_a the minimum description length.

8. Model the intensity profile on scanline s_0 for hypothesis h_2 as a polynomial. Again, use the MDL principle to find the best order and assign its MDL to M_b.

9. Model the scaled intensity profile of scanline s_0 for both h1 and h2 as a polynomial, find the best order using MDL, and assign the smallest MDL to M_c.

10. Compare $(M_a + M_b)$ to M_c. To normalize the results of this test to the range [0,1], we use the measure of merit given by (3),

$$'m = 1 - \frac{M_c - (M_a + M_b)}{M_c + (M_a + M_b)} \tag{3}$$

and any result >1.0 gets set to 1.0.

To obtain a robust measure for a region pair, we average the result of this procedure over all scanlines containing a border pixel, looking either vertically or horizontally depending upon the local border tangent. We then compare this average to the median likelihood and take the more extreme value (towards 0 or 1 depending on whether the average is less than or greater than 0.5, respectively). For more discussion of the profile analysis, see [5].

2.16. Creating the Hypothesis Graph

Once all possible hypothesis pairs are analyzed we generate a hypothesis graph in which each node is a hypothesis and edges connect all hypotheses that are adjacent in the image. We then assign to each edge the likelihood--between 0 and 1--that the two hypotheses it connects are part of the same object. We use the results of the analysis tests to assign weights to edges that represent compatible hypotheses. An example hypothesis graph for the image of two-spheres is shown in Figure 10. All other edges have a weight of 0.0, indicating that they should not be merged in any segmentation.
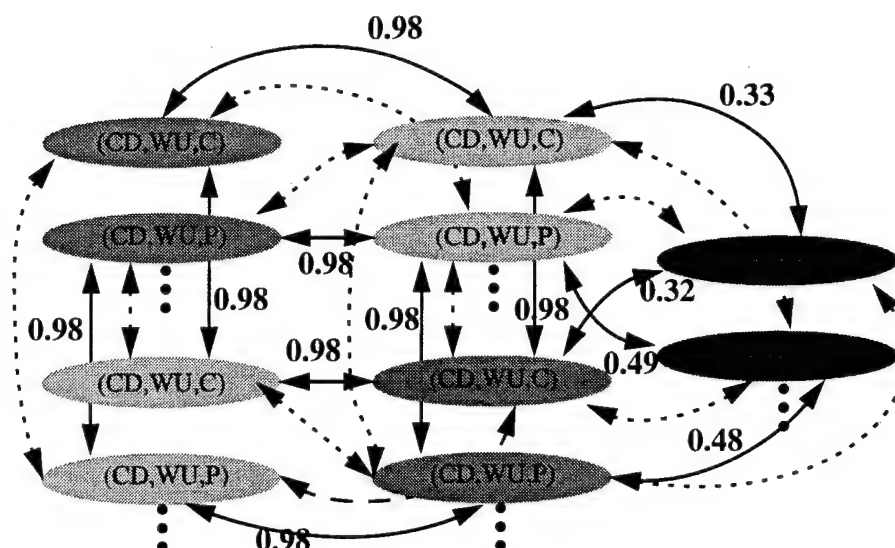
Figure 10: Hypothesis graph for the synthetic test image shown in above right. The numbered solid edges indicate the likelihoods of merging adjacent regions. Dashed edges indicate incompatible hypotheses with a likelihood of 0. All adjacent hypotheses have a "not-merge" edge (not shown) with likelihood 0.5. Note, as more hypotheses are included, the hypothesis graph simply gets more levels.

How best to combine the results of different tests is still an open question. For our current implementation we use a weighted average of the results of the three tests--reflectance ratio, gradient direction, and profile analysis--to get the likelihood of a merger.

Note, however, that each edge actually has two weights associated with it. The weight assigned to the edge is a likelihood that the two hypotheses are part of the same object and should be merged in a segmentation. There always exists the alternative that the two hypotheses are not part of the same object and should not be merged in a segmentation. In order to find "good" segmentations, we must somehow assign a weight to the not-merge alternative. We select a value of 0.5 as the cost of not merging two hypotheses. This is a logical value for the cost of not-merging because it means that not-merging two hypotheses is better than merging them if their likelihood of merging is less than 0.5. For more discussion of the hypothesis graphs, see [5].

2.17. Extracting Segmentations

Extracting segmentations from a single-layer graph of nodes and probabilities has been accomplished by both LeValle & Hutchinson, and Panjwani & Healey for the segmentation of range images and textured surfaces [18][19]. They used a step-wise optimal algorithm, at each step merging the most likely two nodes until some threshold was reached--either the number of regions or a likelihood threshold.

We use essentially the same algorithm. However, the addition of more layers to the graph adds

some new twists. For a more complete discussion, see [8]. One difference is that there will almost always be more than one "best" segmentation--defined as the segmentation with the maximum sum of the log likelihoods of all of the edges within that segmentation. We would like to be able to identify all of these best segmentations. Furthermore, we would like to find a set of segmentations in which each hypothesis is represented at least once.

Our solution to this problem is to find the best segmentation of N different graphs, where N is the number of hypotheses in the image. The basic idea is to set up each graph by selecting one hypothesis and making that the only hypothesis for its region. This forces the segmentation to include it. This results in N different segmentations, where each hypothesis is guaranteed to be included in at least one segmentation. The most likely segmentation out of this group will contain the best grouping of image regions. Because all discontinuous region pairs have a likelihood of 0.5, however, there will almost always be other equally likely segmentations with the same grouping of regions, but different hypotheses for the individual groups. Figure 11(d), for example, shows the set of final segmentations for the image of two spheres. Note that, in the absence of other information, there are four equally likely final segmentations. However, for this example all four segmentations have the same region groupings.
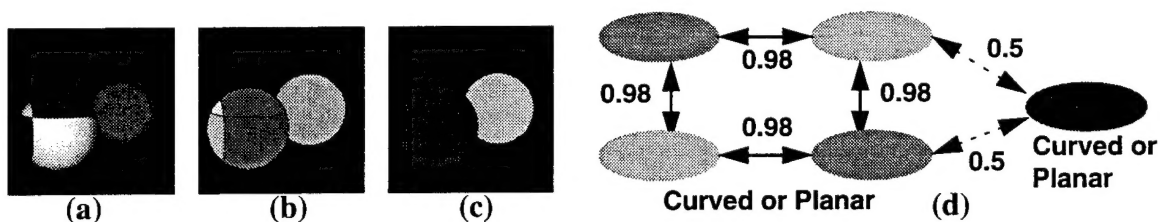

Figure 11: (a) synthetic image of two spheres, (b) initial segmentation, (c) final region grouping of top segmentations, (d) illustration of the hypotheses chosen for the top segmentations. Solid lines indicate merging, dashed lines indicate no merge. The left sphere can be either a planar or curved colored dielectric under white light, as can the right, giving four top final segmentations.

The raw images, initial segmentation, and final segmentation for several test images taken in the Calibrated Imaging Laboratory are shown in Figure 12. The result of our system is a set of region groupings that corresponds more closely to objects in the scene, combined with a high-level description of the form of the illumination and transfer function of those objects. Such a segmentation is much more appropriate for model acquisition and general scene analysis than previous segmentation techniques. Some simple demonstrations of this are given in [8].
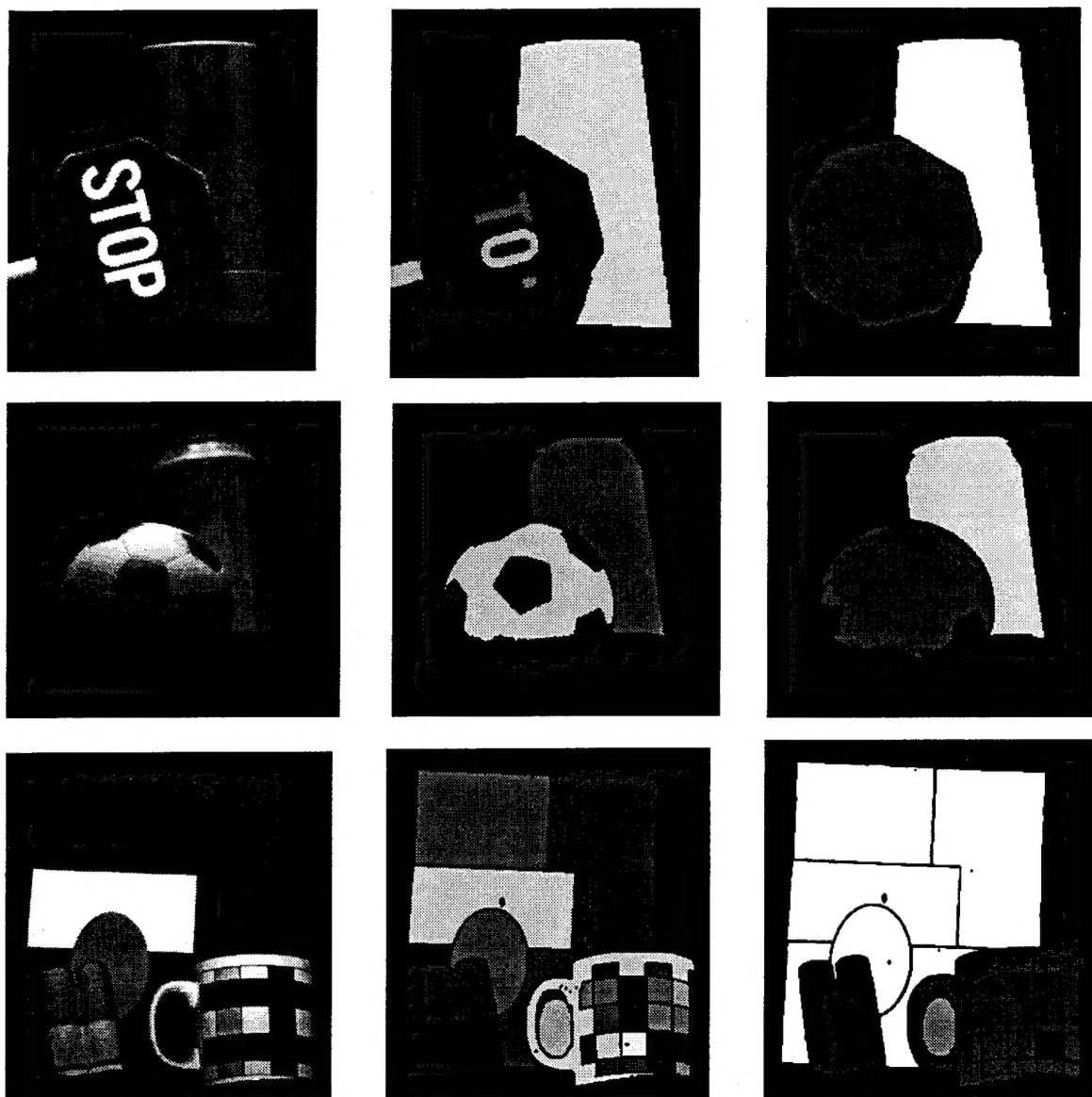
Figure 12: Region groupings of the best segmentations extracted from the hypothesis graphs for some example images taken in the Calibrated Imaging Laboratory. The left image is the raw data, the center image is the initial segmentation, and the right image is the region grouping for the best segmentation extracted from the hypothesis graph.

3. Publications

[1]   M. Maimone, "Computer Vision Home Page," http://www.cs.cmu.edu:8001/afs/cs/user/
cil/vision.html

[2]   B.A. Maxwell and S.A. Shafer, "A Framework for Segmentation Using Physical Models
of Image Formation," CMU Robotics Institute, technical report CMU-RI-TR-93-29.

[3]   B.A. Maxwell and S.A. Shafer, "A Framework for Segmentation Using Physical Models
of Image Formation", *Conference on Computer Vision and Pattern Recognition*, CVPR-
94, IEEE, p. 361-368.

[4]   B. A. Maxwell and S. A. Shafer, "Reasoning About Physical Hypotheses for Scene
Analysis," submitted to *Int'l Conference on Computer Vision*, June 1995.

[5]   B. A. Maxwell and S. A. Shafer, "Physics-based segmentation: Looking beyond color,"
CMU Robotics Institute, technical report CMU-RI-TR-95-37.

[6]   B. A. Maxwell and S. A. Shafer, "Physics-based segmentation: Looking beyond color," in
ARPA Image Understanding Workshop Proceedings, 1995.

[7]   B. A. Maxwell and S. A. Shafer, "Physics-based segmentation: Moving beyond color," to
appear in *IEEE Conference on Computer Vision and Pattern Recognition*, June 1996.

[8]   B. A. Maxwell and S. A. Shafer, "Segmentation for Automatic Model Acquisition and
Scene Analysis," submitted to *Int'l Workshop on Object Representation for Computer
Vision*, April, 1996.

[9]   B. A. Maxwell and S. A. Shafer, "Physics-based Segmentation of Complex Objects Using
Multiple Hypotheses of Image Formation," submitted to *Computer Vision and Image
Understanding*, special issue on Physics-Based Modeling and Reasoning in Computer
Vision, planned for 1996.

4. References

[10]   F.E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis, *Geometrical
Considerations and Nomenclature for Reflectance*, National Bureau of Standards NBS
Monograph 160, Oct. 1977.

[11]   G. J. Klinker, S. A. Shafer and T. Kanade, "Using a Color Reflection Model to Separate
Highlights from Object Color," in *Proceedings of International Conference on Computer
Vision*, IEEE, New York, pp. 145-150, June 1987.

[12]   G. J. Klinker, S. A. Shafer and T. Kanade, "A Physical approach to color image
understanding," *International Journal of Computer Vision* 4(1), pp.7-38, 1990.

[13]   R.C. Bolles, H.H. Baker, and M.J. Hannah, "The JISCT Stereo Evaluation," ARPA Image Understanding Workshop Proceedings, April 1993, p. 263-274.

[14]   M. Bichsel and A.P. Pentland, "A Simple Algorithm for Shape From Shading," in Proc. Conf. Computer Vision and Pattern Recognition, CVPR-92, IEEE, 1992, p. 459-465.

[15]   Q. Zheng and R. Chellappa, "Estimation of Illuminant Direction, Albedo, and Shape From Shading," IEEE Trans. Pattern Analysis and Machine Intelligence, 13(7), 1991, p. 680-702.

[16]   S. K. Nayar and R. M. Bolle, "Reflectance ratio: A photometric invariant for object recognition," in *Proceedings of 4th Int'l Conference on Computer Vision*, Berlin, May 1993, pp. 280-285.

[17]   J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, Singapore, World Scientific Publishing Co. Pte. Ltd., 1989.

[18]   S. M. LaValle, S. A. Hutchinson, "A Bayesian Segmentation Methodology for Parametric Image Models," Technical Report, University of Illinois at Urbana-Champaign Robotics/Computer Vision Series, UIUC-BI-AI-RCV-93-06, 1993.

[19]   D. Panjwani and G. Healey, "Results Using Random Field Models for the Segmentation of Color Images of Natural Scenes," in Proceedings of International Conference on Computer Vision, June 1995, pp.714-719.